ALMA Memo #303
Water Vapour Radiometers for ALMA

Richard Hills and John Richer

Introduction

This is a discussion document setting out the options for performing atmospheric phase corrections by means of radiometry. It is a lightly edited version of a paper prepared for the March 2000 meeting of the Science Advisory Committee. We have tried to take account the comments received but not yet responding to the conclusions of that meeting.

A great deal already been written on this subject. In particular, the relevant memos and other documents have been summarised on the ALMA web site at http://www.alma.nrao.edu/development/cal_imaging/phasecal.html.

Status of current 183 GHz phase correction experiments

The JCMT-CSO single-baseline interferometer was the first to demonstrate phase correction using the 183 GHz line, using equipment built by Martina Wiedner, Richard Hills and colleagues. Only a limited quantity of data were gathered but the results (ALMA memo 252) were encouraging and suggested that even an uncooled system could provide effective phase calibration at submillimetre wavelengths. Single baseline interferometry at JCMT-CSO is no longer a supported mode of operation, so further observations would be difficult though not perhaps impossible to arrange. It is however possible that two SMA antennas can be equipped with 183 GHz systems, using the radiometer currently at CSO plus a clone of it being built in Canada by Christine Wilson and the HIA. It is unclear when this experiment might produce results on Mauna Kea, but access to a large set of data in a variety of atmospheric conditions would certainly be useful in establishing the capabilities of the technique.

On the Chajnantor site, two further 183 GHz radiometers are in operation; these were built as a collaboration between Onsala and Cambridge and are very similar to the Mauna Kea systems, again using uncooled DSB mixers and three roughly 1-GHz wide filters. These two independent systems are aligned with the twin 11-GHz site testing interferometers, with their beams matched as well as possible using newly designed mirrors. The intention is to see how effectively and for what fraction of the time it is possible to use the 183 GHz systems to correct the 11GHz atmospheric phase measurements. It is possible to estimate the height of the turbulence from the lag between the two 11 GHz phase measurements (which are obtained by looking at different satellites) together with information on wind speed and direction. This will be important in establishing how strongly the quality of radiometric phase correction depends on the turbulent scale height, both in practice and through models.

Initial results for both the lag estimates and radiometric phase correction have been obtained in the past 2 months, although operational difficulties (principally power outages, and the difficulties in performing system upgrades and receiver tests on site) have restricted the quantity of data so far obtained. Work on analysing the existing data and on improving the measurements will continue as a high priority, with the goal of producing a report in about 6 months.

Although the results from these more detailed studies will be needed in order to answer some of the questions, we need to have an initial set of specifications for the ALMA radiometer system and a baseline design for inclusion in the plans and cost estimates. We do in fact have sufficient information to provide much of this information already. The following sections summarise our current thinking on the requirements and the design choices.

Design Considerations for the ALMA water vapour monitors

1) Requirements

The first question to be decided is whether we wish to correct just the phase error in the interferometric signal or whether we should also plan to take out the tilts in the wavefront across the individual dishes which cause pointing errors. (The latter effect is sometimes called anomalous refraction, although it is only anomalous in the sense that it would not occur if the atmosphere were uniform.) The correction of such pointing errors with radiometers was discussed by James Lamb and Dave Woody in MMA Memo 224. The problem has been studied in the context of the LMT/GMT 50-metre project by Luca Olmi. (Radio Science, 35, 275-286, Jan 2000.)

In each case we then need to set detailed requirements. We need to decide the path length error allowed as a function of integration time, weather conditions, zenith angle (z) and change in z. For pointing corrections, we need to set the required accuracy (which should be a term in pointing error budget) again presumably as a function of weather and z.

The rms path error given as the goal in existing documents is 38.5 fs (femtoseconds) which is 11.5 micrometers of path. Note that, at this level, the loss of correlation from this cause is only 5% at 950 GHz and 0.7% at 350 GHz, so this is setting the goal very high. (Compare these to the transmission losses of about 70% and 20% for these same frequencies with 1 mm of water vapour.) No reference is made to whether this figure degrades in less than ideal conditions, but is clear that it can be allowed to without seriously affecting the data. A more realistic goal would be to multiply the above figure by $(1 + w_v)$ where w_v is the amount of water in the path in millimetres.

The time allowed for achieving this accuracy is also not presently specified. We have generally been assuming that this refers to a one-second timescale, but we really need to look more closely at the data to see if we are justified in going as fast as this. (Note that the question of whether the correction is applied to the phase in real time or the data taken with short dump-times and stored for later processing has only a small effect on the radiometer requirements but quite large implications for the software.)

A "systematic (avg)" error of 8.4 fs is also quoted in Larry D'Addario's "Phase Stability Specification Note". This is intended to cover things like drifts and offsets that do not integrate down with time in the way that white noise would. The relevant timescale here is the time between calibrations because anything that varies more slowly than this will be taken out by the observations of the calibrator. We will presumably observe calibration sources much less often than we would if we were using only them to remove atmospheric phase fluctuations, but a typical calibration cycle of 50 or 100 seconds seems reasonable. We can, if necessary, move further and use brighter sources than is planned for fast-switching phase correction. Presumably the same observations will generally be used to check the pointing and/or the amplitude calibration. We should try to extend the stability of the radiometers to at least a few times the calibration cycle so that the results on the calibrators give us an independent estimate of how well the correction is working. Stability over about 5 minutes is therefore the requirement. (Note that this implies that the phase stability of the rest of the system must be maintained for at least this length of time.)

It is however essential that we can measure the atmospheric term accurately as we move from source to calibrator. This is certainly more difficult if there are large changes in the total water in the path and/or ground spillover (although it is only the dish-to-dish differences in these effects that are important). At low elevations it would be beneficial to look for calibration sources that are closer to the target in zenith angle than in azimuth, i.e. to search in an elliptical patch of sky.

The value given above (8.4 fs) is an extremely tight specification: it corresponds to less than a degree of phase at 345 GHz, well below that given for the path stability on the antennas for example. We think it unrealistic to insist on this stability from the radiometer system and suggest instead that the value proposed for the short-term noise should apply for the stability over the calibration cycle as well. Obviously the goal will be to do better.

The key sensitivity number is that at the optimum frequency the change in brightness temperature is $\sim 15/w_v$ mK per micrometer of added path. This suggests that a radiometric precision of order 150 mK (corresponding to ~10 microns of path) would be sufficient in good conditions. Given bandwidths of >100 MHz and an integration time of 1 second, this looks reasonable, even for a room temperature mixer, for which Tsys of 1500K should be possible.

For antenna pointing corrections a suitable budget allocation is 0.3 arcsec rms (in dry conditions). This is a wavefront slope of 1.5 microns per metre, which leads to a figure of 9 microns when taken between two points 6 metres apart on the dish. The measurement is however now a difference between two numbers and it probably has to be measured in shorter times than the interferometric phase. This looks marginal with a single uncooled mixer.

Studies of the existing site data (e.g. MMA Memo 223 and references therein) show that much of the observing time will be seriously affected by single dish pointing errors: the overall median seeing is about 1 arcsec compared to the specification for the antennas of 0.6 arc seconds. More study is needed of how fast the pointing fluctuates and how the bad seeing correlates with the other conditions. The obvious conclusion at this stage is that we do need to correct the pointing and that we should assume that this needs to be updated once per second. (With a wind speed of several metres per second and a 12-metre aperture, we can obviously expect some pointing changes on timescales as short as this, but the bulk of the power will normally be a periods of more like 10 seconds.) Note that this has to be done in real time and that we will therefore need to use an algorithm that anticipates the error for a time about one second ahead of the most recent reading.

Other requirements: Compatibility of interfaces (CANbus, etc.) Minimum interference with other systems. A special problem is leakage of the LO and its harmonics into other systems via various paths e.g. out of the feed and by reflection off the subreflector. It is unlikely that we can suppress these completely. The LO's should therefore be locked to system clock so any interference is at an accurately defined frequency. The design should use the fixed reference frequencies already provided at each antenna. (Using 2 GHz and 125 MHz would provide a satisfactory combination). We might add a requirement that the LO can be shifted by a small amount (say 125 MHz) so that any interference can be moved away from a critical line (by about 200 km/s in that case).

Suggested baseline spec: $10(1 + w_v)$ microns of path and $0.3(1 + w_v)$ arc sec of pointing over a 5 minutes of time and 1 degree in z, with 1 sec time resolution.

2) Basic technical approach.

The obvious options are line measurements at 183 GHz, 22 GHz, and in the mid-infrared (10 or 20 microns), or measurement of the (sub)mm continuum as for example used at IRAM.

The latter is unlikely to provide accurate enough path estimates and could not easily accommodate a wide range of conditions.

22 GHz is now essentially ruled out by the size of the optics. The feed would be ~250 mm diameter to measure the interferometric phase and at least 500 mm for correcting the pointing. Sensitivity would in any case be problematical - a cooled system would certainly be required.

The use of infra-red radiometers is a new suggestion from Dr David Naylor (Lethbridge, Canada). The principle is essentially the same as with the millimetre radiometers but uses water vapour emission bands in the mid-IR. The system uses detectors cooled to 77 K. We could not use the telescope’s optics so to measure the pointing corrections we would probably need either several detectors per dish or some optical relay system to give an appropriate spreading of the beam. The initial report on sensitivity and stability looks encouraging, but questions such as how much the results are affected by the temperature and pressure in the fluctuating layer and the effects of cirrus clouds have yet to be investigated. This needs to be done before we can judge whether this might be a viable option for ALMA.

Meanwhile the baseline should remain 183 GHz.

3) Mixer or HFET?

183 GHz HFETS will probably be available but will be expensive, noisy and with poor short-term stability. The baseline should be to use mixers.

4) Cooled or uncooled?

The main advantages of cooled systems are sensitivity and stability. It would also be easy to provide a cold reference load. There is however some concern about how one would calibrate out losses in the Dewar window, especially if there is a possibility of getting dirt or water on it. External optics would almost certainly still be required for the pointing system and it might be possible to introduce some additional calibration signal there. With cooled systems, the radiometer will essentially take up one complete slot in a Dewar and the development path will interact strongly with the main receiver programme. It will also take up some of the cooling power budget (IF amps, windows, connections, etc.) and there would be greater likelihood of LO power leakage.

An uncooled system is clearly simpler, and should cost less to develop and build. Uncooled Schottky mixers can be obtained commercially and are robust and stable.

We therefore believe that an uncooled system should be adopted as the baseline. Assuming, however, that the goal of correcting the pointing is confirmed, there is some question as to whether sufficient sensitivity can be obtained with an uncooled system. Until this is established the cooled option should be kept open as the backup.

Digression on cooled systems:

5a) SIS

If we use SIS mixers, these will have to go in the main Dewar and will presumably be based on the ALMA band 5 mixers. Sensitivity is then excellent and stability almost certainly acceptable given a suitable switching scheme. One can argue that no significant development effort on the mixers is required. The standard IF choice is not ideal (1 to 9 GHz would be better), but we could live with it. For example the LO could be at about 180 GHz so that the upper-sideband IF range of 4 to 12 GHz would correspond to line offsets of ~0.7 to 8.7 GHz. The lower sideband would not be used and would have to be rejected at about the 25 dB level. The mixers would provide a certain amount of sideband rejection and this could be enhanced by having a waveguide filter at the input to the mixer, since the operational frequency is fixed. Although there will naturally be strong resistance to giving up one of the astronomical "slots" (or making the Dewar larger and more complicated), this option is sufficiently attractive that it should probably be kept open for the present. A straw-man design for it could be worked up and costed but no development work seems to be needed now.

We should also consider here the possibility of using the astronomical band-5 receiver to do the radiometry. Given the high sensitivity it might be possible to obtain sufficient accuracy from the shape of the line plus perhaps frequency switching, in which case it should not be necessary to compromise the astronomical performance of the receiver by adding additional switching components inside the Dewar. Another option would be to insert a 45-degree polarising grid into the beam when selecting this mode. This would make it possible to use the two polarisation channels as a cross-correlation receiver. This should also provide a way of doing sideband separation. This would of course mean that correction would not be available when using this receiver for astronomy. (Under good conditions, however, it might be possible to do the water vapour measurements with the band-7 receiver using the 325 GHz water line.) Some additional electronics for generating the LO and processing the IF would need to be added. Extra optics would be needed to do the single-dish pointing corrections and these would have to be inserted into the beam to select this mode.

An important additional consideration is that using an SIS mixer should give sufficient sensitivity to provide a correction for the water vapour *emission* when making total power observations with another receiver. One can see that this should be possible from the fact that, for 1mm of precipitable water, the extra emission ΔT_b for a given Δw_v is several times stronger between 181 and 185 GHz than it is at say 345 GHz.

Again these options seem sufficiently attractive that they should be explored in more detail. The interactions with the rest of the system are nevertheless a substantial negative factor. If nothing else we would be compelled to have band 5 available on all antennas from day 1, which may not coincide with the astronomical priority.

5b) Cooled Schottky

The advantage of using a cooled Schottky system is that it could be housed in a separate Dewar with the band 1 receiver (if that is the outcome of other discussions) where it could be cooled to 15 – 20 K. The interactions with the more critical part of the receiver system would then be reduced. It would however probably be necessary to undertake a new development to obtain suitable mixers and we are not clear what performance could be obtained. The IF amplifiers would probably play a major role here and it may again be best to use the ALMA 4 to 12 GHz ones. If we decide to use a Dicke switch (see below) then we would probably need to develop a suitable coolable switch. This option should be considered further if detailed planning for a band 1 Dewar is undertaken.

Finally in this section, we should note that very compact and relatively cheap refrigerators are now available which could cool a simple radiometer to say 70K. Although reliability might be an issue, it may turn out that this is the most cost-effective way of getting the necessary sensitivity if it cannot be obtained with an uncooled system.

6) Form of switching

For an uncooled system, there seems little chance of obtaining ~0.1 K stability with a total power system given a system temperature of at least 1000 K. (Note that we can get some relief because we are observing a line and are to a considerable extent only concerned with the differences between frequencies. We believe that some form of comparison with a load of known temperature will however be necessary.) We should therefore plan to use either a Dicke switch or a continuous-comparison radiometer which takes the difference between the sky temperature and a temperature-controlled load. For the pointing correction we also need to take the differences between different parts of the aperture. Many options are available but we clearly wish to select the simplest, cheapest and most reliable that can do the job.

The most basic option is a single mixer with a Dicke switch operating between the sky and a fixed-temperature load. Ideally this load should be at a temperature close to that of the sky brightness at which one obtains the best sensitivity (around 170 K). A modulated calibration signal would also be injected via a coupler on the input. An alternative to injecting a cal signal is to switch between the sky and two loads at different temperatures. This gives more flexibility in the choice of temperatures: something like 100 K and 250 K (spanning the sky brightness range of interest) would be best, but combinations like 200 K and 370 K would also be good. The existing MRAO design uses two loads and an optical switching scheme (a “flip-mirror”). This works quite well, but for ALMA it would probably be worth developing an all-electronic switching scheme, using ferrites or diode switches, for both reliability and stability reasons. With a single mixer the system would normally run in double-sideband mode and, provided the gain stability was adequate, the sensitivity would be given by the normal radiometer equation: $\Delta T = 2\ T_{sys}\,(DSB)$ / root(Bt).

The next level of sophistication is to use two mixers. With a hybrid before the mixers and a correlating backend one can then arrange that the output is the difference between the sky temperature and the load. (The use of a correlation receiver in this application is suggested in Luca Olmi’s paper and he refers to the work of Predmore et al. (IEEE Trans MTT-33, 44, 1995) as a successful example of a millimetre-wave continuous comparison radiometer. The sensitivity improves by root 2 and with appropriate switching we can presumably separate the sidebands as well, although the advantages of doing this do not seem very great. (It would perhaps give better information about any contribution from clouds.)

To obtain the gradient in the emission, which gives us the pointing correction, we need to arrange the optics so that the radiometer illuminates a patch on the subreflector, covering about half of it. For a switching scheme the beam then has to be moved around (most naturally as a circular scan about half way out) and the signal put through a pair of synchronous detectors to generate the required error signal. Lamb and Woody suggested a rotating prism to do this but a rotating mirror with its normal slightly tilted with respect to the axis of rotation would also do the job.

An alternative is to again use correlation (i.e. continuous differencing) receivers. The most obvious arrangement would be to have 4 horns in a square, which are optically re-imaged onto the secondary. The two diagonal pairs are connected to 4 mixers via hybrids in such a way that the outputs are the differences in the sky brightnesses required. A mechanism for switching against loads would still be needed to give the interferometric phase correction. Although these schemes sound complicated, the technology does probably now exist to build such combinations of splitters, hybrids and mixers in a stripline form at these frequencies.

More discussion of these schemes seems appropriate before a choice is made here.

7) Form of backend

In principle we could scan the LO and use a fixed and very simple IF with just one fixed frequency. Given that we are struggling for sensitivity this seems unattractive. The stability would probably not be good either. We therefore need a multichannel backend. The obvious choices are a set of filters (as in the MRAO and Onsala systems) and an analogue correlator along the lines developed by Andy Harris. This latter approach is being adopted for a 22 GHz water vapour phase correction scheme on BIMA. (See http://bima.astro.umd.edu/memo/memo67.ps)

More modelling is needed to determine the number of filters required. Studies such as those performed by Bryan Butler (http://www.nrao.edu/~bbutler/work/nraomemos/VLAwvr.ps) in the context of the VLA need to be carried out for the ALMA circumstances. The existing MRAO/Onsala design uses only 3 but it seems likely that at least 4 would be beneficial to give more information about what is going on in the atmosphere. The bandwidth should increase with increasing offset from the line to give more sensitivity where the water emission is weaker: a possible combination might be 0.5 – 1, 1 – 2, 2 – 4 and 4 – 8 GHz. It is of course possible to make a cross-correlation filter spectrometer to use with a correlation front-end, although twice as many filters are needed in a true multiplication scheme.

The analogue correlator form looks attractive as a compact device suitable for mass production. The existing design is limited to about 4 GHz by the analogue multipliers but faster devices are being worked on. An alternative approach using passive detectors is under development at MRAO for CMB work. Because the frequency spacing is fixed, one would need at least 16 lags to cover plus and minus 8 GHz of IF with adequate resolution. (The BIMA scheme expects to use 32 channels.)

We suggest that the analogue correlator be adopted for further investigation with filters as a safe fallback.

8) Local Oscillator

In order to use DSB systems (or a SSB one with modest rejection) we need to put the LO at the line frequency, 183.31 GHz. First harmonic mixers would require 91.155 GHz, which is easy with a fixed-tuned Gunn. Alternatively it may be more economical to adapt components from the standard ALMA LO system even though the tuning flexibility and pahse stability are not required. Fundamental mode mixers are better because there would be fewer LO harmonics and somewhat lower noise. These could be driven with a Gunn plus a doubler, but would need quite a lot of power, especially for several mixers. Using biased mixers rather than self-biasing ones would reduce this problem.

No tuning is needed, except possibly a step of a few MHz to move it out of the way of a particular line. Although with an SSB system one can in principle fit for the frequency, phase locking the LO to the system clock is clearly advisable, so that all the interference spikes are at accurately known frequencies (and with zero fringe rate).

9) Beam Offsets

It is clearly important that the radiometer samples the same path through the atmosphere as the incoming astronomical signal. It is in fact not possible for these to match absolutely perfectly. (For one thing the radiometer signal is incoherent emission from the water molecules and is therefore sampled by the intensity pattern of the antenna, which is always positive. The path length change is a coherent effect and therefore depends on the amplitude pattern. Molecules in certain locations will not contribute to the phase delay and some will even produce an advance!) The question of how well the beams need to overlap depends on how much small-scale structure there is in the water vapour and how far away it is in front of the aperture. We need more data on the height of the fluctuating layers to make quantitative statements on this.

It is however clear that it is desirable to keep the radiometer close to the astronomical feeds but this is not likely to be a very critical parameter because most of the phase fluctuation is in scale that are considerably larger than the beam. If we can place the radiometer feed in the centre of the ring or cluster of feeds, then the beam offsets are likely to be in the range 3 to 10 arc minutes. This corresponds to distances of 1 to 3 metres at a distance of 1 km, i.e. a modest offset compared to the dish diameter. To illuminate a suitable area on the subreflector to be able to do the pointing correction would require a feed about 75 mm in diameter. It is more likely that a much smaller feed (or group of feeds) would be used which would be reimaged onto the subreflector by an optical relay. The final mirror of this could then be in the central position and it would be advisable to allow about 100 mm clear diameter to accommodate it.

The baseline should be to keep the radiometer beam within 10 arc minutes of the astronomical ones and, if it is practical to do so, make this offset smaller than that for the higher frequency channels.

Conclusions

The critical issue at this stage is to decide whether we should aim to correct the single-dish pointing errors or not. Once that is determined more detailed specifications can be drawn up and design choices made. It is also important for the SAC to consider the issue of whether options involving use of the astronomical receivers should be kept open or ruled out now as an undesirable approach.